

USABILITY

Human Research Program - Space Human Factors & Habitability
Space Human Factors Engineering Project
Anikó Sándor, Ph.D., Kritina L. Holden, Ph.D.

ABSTRACT

The Usability project addresses the need for research in the area of metrics and methodologies used in hardware and software usability testing in order to define quantifiable and verifiable usability requirements. A usability test is a human-in-the-loop evaluation where a participant works through a realistic set of representative tasks using the hardware/software under investigation. The purpose of this research is to define metrics and methodologies for measuring and verifying usability in the aerospace domain in accordance with FY09 focus on *errors, consistency, and mobility/maneuverability*.

Usability metrics must be predictive of success with the interfaces, must be easy to obtain and/or calculate, and must meet the intent of current Human Systems Integration Requirements (HSIR). Methodologies must work within the constraints of the aerospace domain, be cost and time efficient, and be able to be applied without extensive specialized training.

SHFE RISKS ASSOCIATED WITH DRP

Risk of Error Due to Inadequate Information. When conducting usability testing, if usability metrics are incompletely or incorrectly defined, the information acquired from user testing may be inadequate and may lead to erroneous design decisions. In turn, those erroneous design decisions can result in a user interface that does not provide the information needed by the crew.

Risk of Reduced Safety and Efficiency Due to Poor Human Factors Design. Good human factors design depends greatly on proper human-in-the-loop testing and appropriate usability metrics. If any of these are inadequate, design flaws will not be identified, and the design will have reduced safety and efficiency.

Gap addressed: Currently, Constellation usability requirements in the Human Systems Integration Requirements (Rev. C) document are defined in terms of errors: minimal impact errors and significant impact errors. While the requirements specify maximum error rates, the details of how to define an error, and how to calculate error rates are not provided. Similarly, there is a gap in knowledge in defining metrics for consistency and mobility/maneuverability, and this DRP will help define them.

INTRODUCTION

- This DRP aims to develop usability metrics that will help formulate verifiable usability requirements.
- Currently, Constellation usability requirements in the Human Systems Integration Requirements (Rev. C) document are defined in terms of errors: minimal impact errors and significant impact errors.
- While the requirements specify maximum error rates, the details of how to define an error, and how to calculate error rates are not provided.

Definition of Usability

The International Standards Organization ISO 9241-11 defines usability as “The extent to which a product can be used by specified users to achieve specified goals”, and recommends evaluating usability in terms of measures of effectiveness, efficiency, and satisfaction.

Measures of effectiveness (i.e. Can you accomplish the task?) relate the goals or sub-goals of the user to the accuracy and completeness with which these goals can be achieved.

Measures of efficiency (i.e. Can you accomplish the task in an ideal timeframe and use of resources?) relate the level of effectiveness achieved to the expenditure of resources.

Satisfaction (i.e. Do you like the system?) measures the extent to which users are free from discomfort, and their attitudes towards the use of the product. ISO 9241-11 also mentions the additional metrics of cognitive and physical workload.

ISO 9126 document on *Software engineering - Product Quality Metrics* describes a Software Quality Model (See Figure 1) that includes usability. Within this model, usability is defined as a quality metric along with functionality, reliability, efficiency, maintainability, and portability.

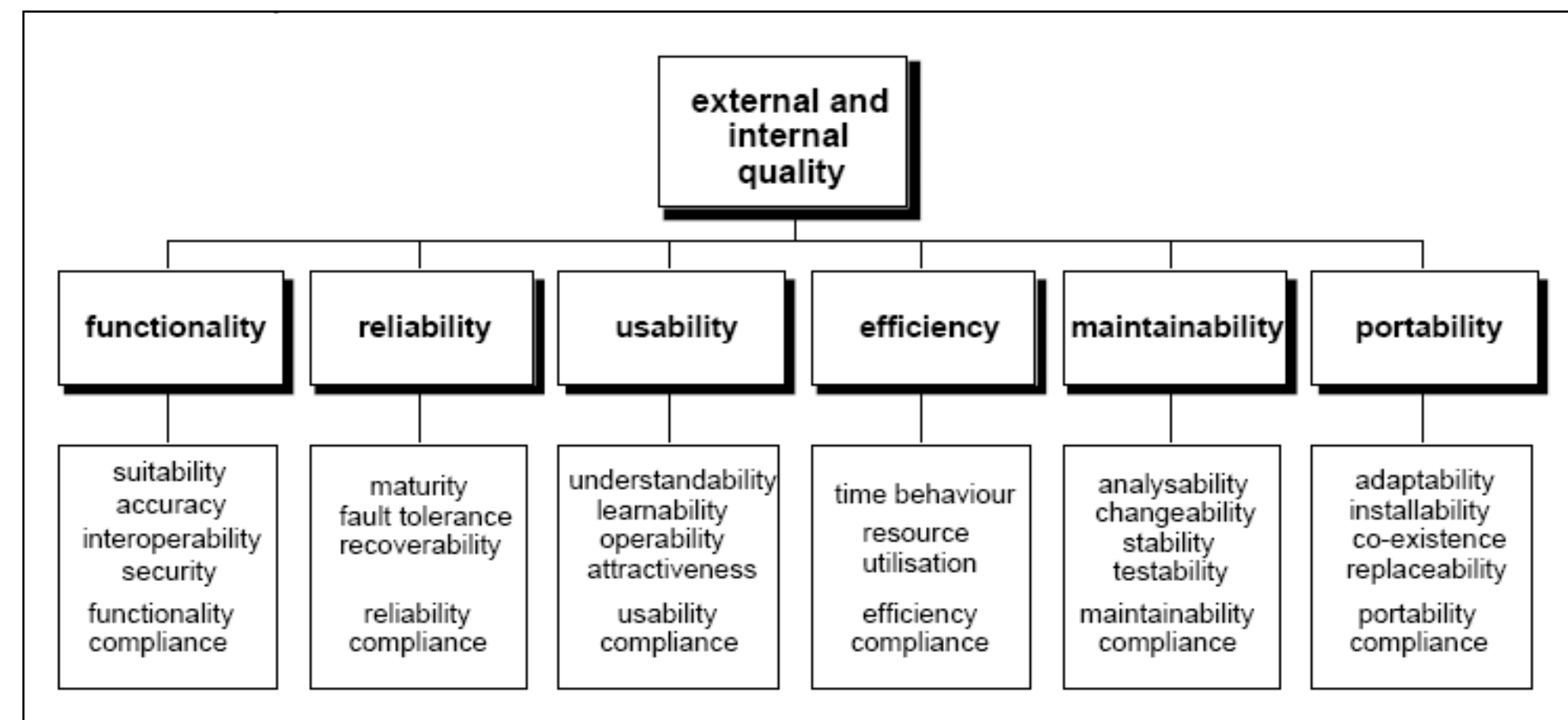


Figure 1. ISO 9126 Software Quality Model

Jacob Nielsen's Definition

Nielsen (1993) describes usability in terms of five factors: learnability, efficiency, memorability, errors, and satisfaction.

Learnability refers to the ease of accomplishing basic tasks when users encounter the design for the first time. Learnability expresses how well a novice user can use the system, while the efficient use of the system by an expert is expressed by efficiency. If the system is used only occasionally, the term *memorability* is used.

Efficiency can be defined as time needed to accomplish the task after users are already familiar with the design.

Errors can be counted during performance observation and rated based on severity.

User satisfaction indicates how pleasant is to use the design.

USABILITY TESTING METHODOLOGY

Human Centered Design

- Human Centered Design* (HCD) is an approach (See Figure 2) that focuses on making a system usable by incorporating human factors and ergonomics in system design (ISO 13407).
- HCD is characterized by early and frequent user involvement and an iterative design-test-redesign process. Usability testing is one of the key methods within the HCD approach.

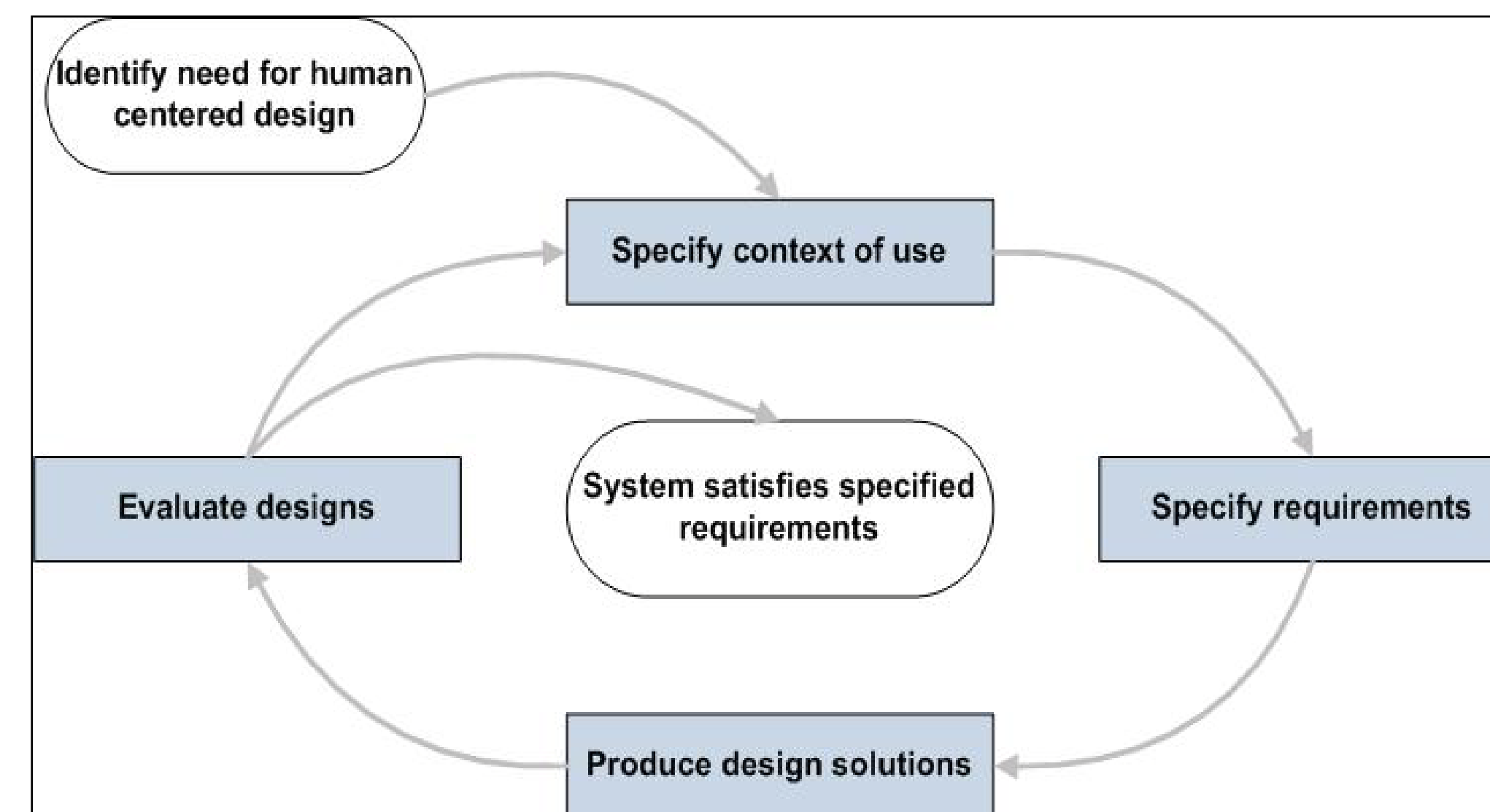


Figure 2. Human Centered Design process model

Tasks

- Relevant tasks have to be selected for the hardware or software to be tested. These tasks may be defined based on task analysis or based on the focus of the usability test. Standard practice is to select several types of tasks for testing:
 - tasks that are frequent and nominal,
 - tasks that are difficult or expected to cause problems, and
 - tasks that are off-nominal or rarely performed.
- Based on the tasks, the test conductor constructs realistic scenarios that are presented to the participant. For example, one such scenario in the context of an online word processor application may be the following:
 - Log in to the website using your username and password.
 - Create a new document with the title “My document”.
 - Save the document and close it.
 - Change the name of the document to “My first document”.

Selection of Participants and Sample Size

- It is recommended to select participants who are representative of the user group of the software or hardware in question. Sample size is usually decided based on availability of participants and cost; however, it is recommended to have at least 10, if possible, 20 or more participants to make sure that even usability problems with lower probability are found during testing.
- Usability testing can be used to compare designs or products and it can be used also for verification purposes. However, for the latter case, one has to define the success criteria for the software or hardware in terms of the metrics that have been used during the testing phase, or that have been mandated in requirements.

Defining the context of usability testing

Systems should be tested in a context as similar as possible to that of the actual system, and results should be interpreted in the light of the context. For example, if a system is used under high stress, results from a laboratory evaluation that is low stress must be interpreted with caution. Results can sometimes be extrapolated by assuming that error rates will be higher under stress, and also that task times will change.

USABILITY METRICS

Metrics of Interest for FY09

Errors

- Before conducting usability testing, the researcher must decide on the definitions of errors and on definitions of severity levels (Tullis and Albert, 2008). A very strict definition of errors could include number of comments or statements about confusing interface elements (for example “I am not sure which button to click”) or longer response times. A more lenient definition might consider only erroneous clicks, or an inability to complete the task as errors. Currently, Constellation usability requirements in the Human Systems Integration Requirements (Rev. C) document are defined in terms of errors: minimal impact errors and significant impact errors. Although the requirements specify maximum error rates, the details of how to define an error, and how to calculate error rates are not provided.
- Errors are one of the standard accepted metrics employed in usability testing - far more complex than may appear on the surface.
- Some specific questions to be addressed in this DRP with respect to errors are:
 - How can errors be defined and classified?
 - What is a usability error versus a human error?
 - How are errors measured? What about recoverable errors?
 - How is error severity taken into consideration?
 - How are usability errors related to risk assessment?

Readability and Legibility

- Readability and legibility are important aspects of interface usability
- This DRP will provide a standard methodology for readability/legibility measurements to help verify requirements for readability/legibility

Consistency

- Consistency is the unification of the general operating sequence, terminology, components, layout, color, and style in an application (Shneiderman, 1998).
- In a consistent interface, if one part of the software behaves in a certain way, the other parts will also provide the same type of interaction.
- Ozok and Salvendy (2004) developed a scale using several factors of consistency: text structure, general text features, information presentation, lexical categories, meaning, user knowledge, text content, communication attributes, and physical attributes. However, their guidelines refer to interfaces heavy in text and do not give enough guidance on general consistency.
- The work in this DRP includes development of a consistency scale that is applicable to more graphical user interfaces, and user interfaces in general. The new scale will include categories for:
 - presentation of information to the user;
 - input of information to the system, and
 - method of interaction between the user and the system.

Mobility/Maneuverability

- Even though the Cooper-Harper provides a subjective measure of a person's ability to control and stabilize the hardware, it focuses on stability rather than mobility/maneuverability.
- Objective data (e.g., range of motion or torque) have been used to quantify mobility of space suits; however, the need for a user to subjectively rate the mobility/maneuverability of hardware as a whole, while completing a specific task is critical and not addressed by currently available scales.
- A standardized hardware maneuverability/mobility usability measurement and methodology needs to be developed in order to help practitioners to measure the usability of various types of hardware.